あるSEQUENTIAL STOCHASTIC ASSIGNMENT PROBLEMについて (学習と制御とその周辺)

AUTHOR(S):

Nakai, Toru

数理解析研究所講究録
557巻 1985年 164-187

ある SEQUENTIAL STOCHASTIC ASSIGNMENT PROBLEM について

大阪府立大学総合科学部 中井 達 (Tōru Nakai)

1. Introduction

In an interesting paper [3] by Derman, Lieberman and Ross, a sequential stochastic assignment problem is investigated. Here we treat this problem in a partially observable Markov chain with a known transition probability matrix. Unlike problems treated in Derman, Lieberman and Ross [3], Nakai [6], [8], it is assumed that the states of this chain are not observable, however an a priori probability distribution p of the states is given.

A non-negative random variable is associated with each state of the process, and the precise relationship between the states and the random variables is previously known. The decision-maker observes a realization of these random variables one by one sequentially. The number of actions available to the decision-maker is finite, and each action is used only once. After observing a realization x of the random variable associated with the current state of the process, the decision-maker updates information in the Bayesian manner, and selects one of N actions $\{ a_1,\ldots,a_N \}$. If he selects the i-th action a_i, then he assigns this action to x, earns a reward of $a_i u(p,x)$ (where u(p,x) is an appropriate reward function), and moves to a next state of the chain, at the next instant, where he will have only the remaining N - 1 actions to choose from. When

all the N actions are used up the process stops. The objective of this problem is to choose one of N! permutations of the N available actions so as to maximize the total expected reward.

In a formerly studied sequential stochastic assignment problem, the random variable means a worth of an arriving job, and a worth of each job is distributed as the independently and identically distributed random variables. Here we make a more realistic assumption that the succesive distributions of the random variables are governed by a Markov chain, but the states of this chain are not known explicitly. In this problem, the states correspond to the economic conditions, and a worth of each job depends on this conditions. The learning procedure about these conditions is introduced.

The problem often refered as the house selling problem is a special case of this problem. (Albright [1], Sakaguchi [14]) The problem is one of waiting for the highest bid of price among N objects, and this price depends on the conditions of the demand. The learning procedure may be also important in this problem as the problem treated here.

There is another interpretation of this problem. Concerning the inequalities, the following property is treated as Hardy's theorem. (See Hardy, Littlewood and Polya [4, p. 391])

" If $x_1 \geq x_2 \geq \cdots \geq x_N \geq 0$ and $y_1 \geq y_2 \geq \cdots \geq y_N \geq 0$, then

$$\max_{\sigma \in \mathfrak{S}_N} \sum_{k=1}^{N} x_k y_{\sigma(k)} = \sum_{k=1}^{N} x_k y_k,$$

where $\mathfrak{S}_N$ is the symmetric group on N letters. "

A sequential stochastic assignment problem will be considered as a stochastic generalization of this property. At each time when the decision-maker observes a realization x of the random variable, he must select the best action among the N actions one at a time in sequential order. Here we assume the inequality $a_1 \geq \cdots \geq a_N \geq 0$ without loss of generality, similarly to [3].

In Nakai [8], a problem where the states of the chain are always known to the decision-maker is observed. A problem over an infinite horizon is, however, treated there and an action to take an option to pass is introduced.

In Section 2, several assumptions are introduced. For a set of informations about the states of the chain, we introduce a relation and observe some fundamental properties in Section 3. These things are obtained through a method similar to one used in Nakai [7]. In Section 4, we formulate this problem by means of the dynamic programming. The theorem which contains an optimal policy and the total expected reward under this policy, is treated in Section 5. As concerns the sequential stochastic assignment problem with N actions in [3], etc, the same critical number policy, which is determined by the (N-1) critical numbers, is optimal for any positive values of $a_1,\ldots,a_N$. On the other hand, the optimal policy obtained here is not always a critical number policy: we will consider a simple example in Section 5.

A sequential stochastic assignment problem includes a problem of optimal selections, i.e., a problem is to select the k best realizations of the random variables out of N where the reward is the sum of the k values selected. As for the problem treated here, if we put $a_1 = \cdots = a_k = 1$

and $a_{k+1} = \ldots = a_N = 0$ ($k \leq N$), then this is a problem of optimal selections. According to a realization of a random variable associated with each applicant, the decision-maker can select k applicants out of N in order to maximize the total expected reward. When k = 1, this is an optimal stopping problem which is formulated as a partially observable Markov decision process as in Monahan [5]. For this problem of optimal selections, we observe a relation to former results of a sequential stochastic assignment problem.

2. Partially observable Markov chain

We will consider a process which is observed at time points t = 1, 2,..,N to be in one of a number of possible states. The set of possible states is to be assumed countable and will be labelled by the positive integers 1, 2,.... Let $\{ Y_t, t = 1,\ldots,N \}$ be the above stationary Markov chain with a known transition probability matrix $P = (p_{ij})$ (i, j = 1,2,..). We assume that the states of the process are not observable, but a priori information about the states is given. All information is summarized by a probability distribution p on { 1, 2,... } ($p \in S$ and $S = \{ p \mid p = (p_1, p_2, \ldots.), p_i \geq 0$ and $\sum_{j=1}^{\infty} p_j = 1 \}$).

A non-negative random variable X_i is associated with each state i of the process, and the X's are independent of the other states and the time points. The probability distribution function $F_i(x)$ of X_i is assumed to be absolutely continuous with a density function $f_i(x)$. (i = 1,2,....)

Similarly to the problem treated in Nakai [7], the following assumptions are introduced.

Assumption 1. When the process is in state i, i.e., $Y_t = i$ ($i = 1,\ldots,N$), the conditional expectation of X_i is finite and bounded in i. The density function $f_i(x)$ is uniformly bounded in i.

Assumption 2. If $j < i$ ($i, j = 1,2,\ldots$), $f_j(x)f_i(y) \leq f_i(x)f_j(y)$ for any x and y ($x \leq y$), i.e., $f_j(x)/f_i(x)$ is non-decreasing in x. ($x \in \{ x \mid f_i(x) \neq 0 \}$)

Assumption 3. If $1 \leq m < k$, there exists $x_{mk} = \sup \{ x \mid f_m(x) \leq f_k(x) \}$, which satisfies the following inequalities.

If $x \geq x_{mk}$, then $f_m(x) > f_k(x) > f_{k+1}(x) > \ldots\ldots$,

or otherwise $f_k(x) \geq f_m(x) \geq f_{m-1}(x) \geq \ldots\ldots \geq f_1(x)$.

Assumption 4. If $j < i$ ($i, j = 1,2,\ldots$), then

$p_{ki}p_{mj} \geq p_{kj}p_{mi}$ for any m and k ($m < k$).

First three assumptions are satisfied for a sequence of the exponential with densities $f_i(x) = \lambda_i \exp(-\lambda_i x)$ ($\lambda_1 \leq \lambda_2 \leq \ldots\ldots$).

The likelihood ratio ordering is introduced in Assumption 2, see Ross [13, p. 266]. If this process is a Markov chain with two states, Assumption 4 is equivalent to the inequality that $p_{11} \geq p_{21}$ ($p_{12} \leq p_{22}$), which is a well known assumption stated in Ross [12], Monahan [5]. From Assumption 4, this Markov chain is total positive of order two.

After observing a realization x of the random variable associated with the current state of the process, the decision-maker updates information in the Bayesian manner. It is assumed that, for any realization x and

probability distribution p in S, the posterior probability distribution exists and is specified by the Bayes' theorem. Let $T(p,x) = (T_1(p,x), T_2(p,x), \dots)$ be the posterior probability distribution for any $x \in R_+ = [0, \infty)$ and $p \in S$; then

$$T_i(p,x) = p_i f_i(x) / (\sum_{j=1}^{\infty} p_j f_j(x)). \quad (i = 1,2,\dots) \tag{1}$$

After improving this information, the process will make a transition according to the transition probability matrix P, and, at the next instant, information about the states will be $\bar{T}(p,x) = T(p,x) \cdot P$, where

$$\bar{T}(p,x) = (\bar{T}_1(p,x), \bar{T}_2(p,x), \dots)$$

and

$$\bar{T}_j(p,x) = \sum_{i=1}^{\infty} T_i(p,x) p_{ij}. \quad (p \in S, x \in R_+ \text{ and } j = 1,2,\dots) \tag{2}$$

3. Partial order in S

Similarly to Nakai [7], we introdece a relation in S by the following manner.

Definition 1. For p and q (p, q $\in$ S), $p > q$ if $p_i q_j \leq p_j q_i$ for any i and j (i > j and i, j = 1,2,...) and $p_i q_j < p_j q_i$ for at least one pair of values i and j. If $p_i = q_i$ (i = 1,2,...), p = q. $p \geq q$ if and only if p > q or p = q.

For this relation, we obtain the following properties which are showed by a method similar to one used in Nakai [7] and we omit the proofs here.

Lemma 1. The relation defined by Definition 1 is a partial order, and $p^* = (1,0,0,\ldots) \geq p$ for any p in S.

Throughout this paper, we call a real valued function u(p) on S is non-decreasing in p if and only if $u(p) \geq u(q)$ for $p \geq q$ ($p, q \in S$).

Lemma 2. If $u(p) = \sum_{i=1}^{\infty} a_i v_i(p) p_i$ ($a_1 \geq a_2 \geq \ldots\ldots$, $v_1(p) \geq v_2(p) \geq \ldots\ldots$ and $v_i(p)$ is increasing in p), then u(p) is increasing in p.

Lemma 3. If $x \geq y$, then $\bar{T}(p,x) \geq \bar{T}(p,y)$ for any p in S.

Lemma 4. If $p \geq q$ ($p, q \in S$), $\bar{T}(p,x) \geq \bar{T}(q,x)$ for any $x \geq 0$.

Without Assumptions 1 - 4, Lemmas 1 - 4 are not obtained in a general case.

4. Formulation of the problem

When information about the state of the process is p ($\in S$) and N actions { $a_1,\ldots,a_N$ } are available to perform, we call this problem as in state $(a_1,\ldots,a_N;p)$. We consider that a non-negative number a_i is associated with each action ($i = 1,\ldots,N$), and assume that $a_1 \geq \ldots \geq a_N$ without loss of generality.

When the process is in state $(a_1,\ldots,a_N;p)$, after observing a realization x of the random variable associated with the current state of the process, the decision-maker updates information as T(p,x) and selects one of N available actions. If he selects the i-th action a_i, he earns an immediate reward of $a_i \cdot w(T(p,x),x)$ (where w(p,x) is an appropriate

reward function), and the selected action is unavailable for future decisions. After the decision, the process will next make a transition according to the transition probability matrix P which is not affected by actions. At the next instant, the problem is in state $(a_1,..,a_{i-1},a_{i+1},..,a_N;\bar{T}(p,x))$, and we then face a problem equivalent to one that starts in state $(a_1,..,a_{i-1},a_{i+1},..,a_N;\bar{T}(p,x))$.

$w(p,x)$ is a non-negative function defined on $S \times R_+$. It is assumed that $w(p,x)$ is increasing in x and p in the sense of Section 3. Moreover $w(p,x)$ is measurable and $\int_0^\infty w(p^*,x)dF_1(x) < \infty$. If the decision-maker earns a reward $w_i(x)$ when the process is in state i (i = 1,2,...) and $w_i(x) \geq 0$, $\uparrow$ x and $\downarrow$ i with $\int_0^\infty w_1(x)dF_1(x) < \infty$, then $w(p,x) = \sum_{i=1}^\infty p_i w_i(x)$ satisfies the above conditions. If we put $u(p,x) = w(\bar{T}(p,x),x)$, then $u(p,x)$ also satisfies the above conditions by Lemmas 3 and 4. In the following discussions, we use the notation $u(p,x)$ for convenience sake.

Throughout this paper, we assume that N is equal to the number n of actions available for the assignment. This restriction can be relaxed easily: if n < N, add N - n actions having a's equal to zero associated with them, and if n > N, we can disregard the n - N inefficient actions. Similar restrictions are found in [3], etc.

The policy is to choose one of N! permutations of N available actions for a sequence of N random variables; at each time when the decision-maker observes a realization x of the random variable X, he must select the best action among the N actions one at a time in sequential order. Whenever the problem is in state $(a_1,\ldots,a_N;p)$, this sequential stochastic assignment problem is called by $P_N(a_1,\ldots,a_N;p)$, and the total expected reward obtainable under an optimal policy is denoted by $v_N(a_1,\ldots,a_N;p)$.

This problem is modelled by an argument similar to one used in a Markov decision process, (see Ross [10, Chapter 6]) and, therefore, from a dynamic programming formulation, $v_N(a_1,\ldots,a_N;p)$ satisfies the following recursive equations.

$$v_N(a_1,\ldots,a_N;p) = \int_0^\infty v_N(a_1,\ldots,a_N;p|x)dF_p(x), \tag{3}$$

and

$$v_N(a_1,\ldots,a_N;p|x) = \max_{1\le i\le N}\{ a_i u(p,x) + v_{N-1}(a_1,..,a_{i-1},a_{i+1},..,a_N;\bar{T}(p,x)) \}, \tag{4}$$

with $v_1(a_1;p|x) = a_1 u(p,x)$, where $F_p(x) = \sum_{j=1}^\infty p_j F_j(x)$.

5. Main result

First we construct a sequence of functions and observe several properties about this sequence. The theorem which contains an optimal policy and the total expected reward under this policy is derived from these properties.

Let $g(x)$ and $u(x)$ be non-negative measurable functions. When $\int_0^\infty g(x)dF(x)$ and $\int_0^\infty u(x)dF(x)$ exist, let $U_F(u(x),g(x))$ and $V_F(u(x),g(x))$ be functions such that

$$U_F(u(x),g(x)) = \int_0^\infty (u(x) - g(x))^+ dF(x) \tag{5}$$

and

$$V_F(u(x),g(x)) = \int_0^\infty g(x)dF(x) + U_F(u(x),g(x)), \tag{6}$$

where $F(x)$ is an absolutely continuous probability distribution function defined on R_+ and $h(x)^+ = \max\{ h(x),0 \}$. These non-negative functions

are considered as generalizations of well known functions, $T_F(z)$ and $S_F(z)$, in DeGroot [2, p. 246]. If $u(x) = x$ and $g(x) = z$, then

$$U_F(u(x),g(x)) = T_F(z) \text{ and } V_F(u(x),g(x)) = S_F(z).$$

Let $\{ g_{N,i}(p) \}$ ($p \in S$ and $1 \leq i \leq N$) be a sequence of functions defined by the following manner.

$$g_{N,i}(p) = V_{F_p}(u(p,x),g_{N-1,i}(\bar{T}(p,x))) - U_{F_p}(u(p,x),g_{N-1,i-1}(\bar{T}(p,x))) \quad (7)$$

and

$$g_{N,0}(p) = \infty, \; g_{N,N+1}(p) = 0 \; (N \geq 0).$$

In the following discussions, the next notations will be used. Let

$$S_{N,i}(p) = \{ x \mid g_{N-1,i}(\bar{T}(p,x)) \leq u(p,x) < g_{N-1,i-1}(\bar{T}(p,x)) \}, \quad (8)$$

$$U_{N,i}(p) = \bigcup_{j=1}^{i-1} S_{N,j}(p),$$

and

$$L_{N,i}(p) = R_+ - U_{N,i+1}(p)$$

where $U_{N,1}(p) = L_{N,N}(p) = \emptyset$ and $U_{N,N+1}(p) = R_+$.

Since $g_{1,1}(p) = \sum_{j=1}^{\infty} p_j \int_0^{\infty} u(p,x) dF_j(x)$, the sequence defined by (7) is well defined. Concerning these functions, we have the following properties. First we state these things and put the proofs together.

Proposition 1. $g_{N,i}(p)$ is increasing in p for any N and i.

Proposition 2. $g_{N,i}(p)$ is decreasing in i for any p and N.

Proposition 3. $g_{N,i}(p)$ is increasing in N for any p and i.

Corollary 1. The sets, $S_{N+1,i}(p)$, $U_{N+1,i}(p)$ and $L_{N+1,i}(p)$, are disjoint with each other and

$$S_{N+1,i}(p) \cup U_{N+1,i}(p) \cup L_{N+1,i}(p) = R_+.$$

Corollary 2. Let $h_{N+1,i}(p|x)$ be a function defined by

$$h_{N+1,i}(p|x) = g_{N,i-1}(\bar{T}(p,x))I_{U_{N+1,i}(p)} + u(p,x)I_{S_{N+1,i}(p)} + g_{N,i}(\bar{T}(p,x))I_{L_{N+1,i}(p)}, \tag{9}$$

where I_A is an indicator function of the set A. Then we have

$$g_{N+1,i}(p) = \int_0^\infty h_{N+1,i}(p|x)dF_p(x), \tag{10}$$

i.e., $h_{N+1,i}(p|x)$ is an integrand of $g_{N+1,i}(p)$.

Proposition 4. $U_{N+1,i}(p) \subset U_{N,i}(p)$ for any p and i. ($1 \leq i \leq N$)

Proposition 5. If $w(p,x) = w(x)$, i.e., $w(p,x)$ and $u(p,x)$ do not depend on p, then $U_{N+1,i}(p) \subset U_{N+1,i}(q)$ for $p \geq q$ ($p, q \in S$) and $1 \leq i \leq N + 1$.

Here we employ the induction principle on N. When N = 1,

$$g_{1,1}(p) = \sum_{j=1}^\infty p_j \int_0^\infty u(p,x)dF_j(x)$$

and $g_{1,0}(p) = \infty$. Since $v_i(p) = \int_0^\infty u(p,x)dF_i(x)$ is increasing in p and $v_j(p) \geq v_i(p)$ ($j \leq i$ and i, j = 1,2,...), Lemma 3 yields Proposition 1, and Proposition 4 is derived from this property. The other properties are obvious by the definition for N = 1.

Next we consider the general case.

Proof of Proposition 1. Lemma 4 and the induction assumption yield that $g_{N-1,i-1}(\bar{T}(p,x))$ and $g_{N-1,i}(\bar{T}(p,x))$ are increasing in p. Let's compare two functions $g_{N,i}(p)$ and $g_{N,i}(q)$. ($p \geq q$)

If

$$h_{N,i}(p|x) \geq h_{N,i}(q|x) \tag{11}$$

for any x in R_+, this proposition is obtained. From Proposition 4 for N - 1, we compare two functions $h_{N,i}(p|x)$ and $h_{N,i}(q|x)$ in the following nine cases; a) $U_{N,i}(p) \cap U_{N,i}(q)$, b) $U_{N,i}(p) \cap S_{N,i}(q)$, c) $U_{N,i}(p) \cap L_{N,i}(q)$, d) $S_{N,i}(p) \cap U_{N,i}(q)$, e) $S_{N,i}(p) \cap S_{N,i}(q)$, f) $S_{N,i}(p) \cap L_{N,i}(q)$, g) $L_{N,i}(p) \cap U_{N,i}(q)$, h) $L_{N,i}(p) \cap S_{N,i}(q)$ and i) $L_{N,i}(p) \cap U_{N,i}(q)$. These sets are disjoint with each other. It is easy to show Inequality (11); for example, concerning Case d), if $x \in S_{N,i}(p) \cap U_{N,i}(q)$, then

$$g_{N-1,i}(\bar{T}(p,x)) \leq u(p,x) < g_{N-1,i-1}(\bar{T}(p,x))$$

and

$$g_{N-1,i-1}(\bar{T}(q,x)) \leq u(q,x).$$

Since $x \in S_{N,i}(p) \cap U_{N,i}(q)$,

$$h_{N,i}(p|x) = u(p,x) \text{ and } h_{N,i}(q|x) = g_{N-1,i-1}(\bar{T}(q,x)).$$

Therefore the fact that u(p,x) is increasing in p, implies

$$h_{N,i}(p|x) \geq h_{N,i}(q|x).$$

Concerning Case h), if $x \in L_{N,i}(p) \cap S_{N,i}(q)$, then

$$u(p,x) < g_{N-1,i}(\bar{T}(p,x))$$

and

$$g_{N-1,i}(\bar{T}(q,x)) \leq u(q,x) < g_{N-1,i-1}(\bar{T}(q,x)).$$

Moreover we have

$$h_{N,i}(p|x) = g_{N-1,i}(\bar{T}(p,x)) \text{ and } h_{N,i}(q|x) = u(q,x).$$

Since $u(p,x) \geq u(q,x)$,

$$h_{N,i}(p|x) \geq h_{N,i}(q|x).$$

The other cases are obtained similarly.

Proof of Proposition 2. Similarly to the proof of Proposition 1, since $S_{N,i}(p) \cap S_{N,i-1}(p) = \emptyset$, we compare two functions $h_{N,i}(p|x)$ and $h_{N,i-1}(p|x)$ ($p \in S$) in four regions; a) $L_{N,i}(p)$, b) $S_{N,i}(p)$, c) $S_{N,i-1}(p)$ and d) $U_{N,i-1}(p)$. These sets are disjoint with eash other and the union of these sets is equal to R_+.

Concerning Case c), if $x \in S_{N,i-1}(p)$, then

$$g_{N-1,i-1}(\bar{T}(p,x)) \leq u(p,x) < g_{N-1,i-2}(\bar{T}(p,x)).$$

Since $S_{N,i-1}(p) \subset U_{N,i}(p)$, $h_{N,i-1}(p|x) = u(p,x)$ and $h_{N,i}(p|x) = g_{N-1,i-1}(\bar{T}(p,x))$. Therefore we have the inequality

$$h_{n,i}(p|x) \leq h_{N,i-1}(p|x).$$

Concerning the other cases, the inequality is obtained similarly.

Proof of Proposition 3. Similarly to the above propositions, we compare two functions $h_{N,i}(p|x)$ and $h_{N-1,i}(p|x)$ in five regions; i.e., from the induction assumption and Proposiiton 4, a) $L_{N-1,i}(p)$, b) $S_{N-1,i}(p) \cap L_{N,i}(p)$, c) $(S_{N-1,i}(p) \cap S_{N,i}(p)) \cup (U_{N-1,i}(p) \cap L_{N,i}(p))$, d) $U_{N-1,i}(p) \cap S_{N,i}(p)$ and e) $U_{N,i}(p)$. Since

$$g_{N-1,i-1}(\bar{T}(p,x)) \geq g_{N-2,i-1}(\bar{T}(p,x))$$

and $g_{N,i}(\bar{T}(p,x)) \geq g_{N-1,i}(\bar{T}(p,x))$,

the inequality $h_{N,i}(p|x) \geq h_{N-1,i}(p|x)$ is derived from an argument similar to one used in the above propositions. For example, if we consider Case b), then the inequalities

$$g_{N-2,i}(\bar{T}(p,x)) \leq u(p,x) < g_{N-1,i}(\bar{T}(p,x))$$

are realized. Since

$$h_{N,i}(p|x) = g_{N-1,i}(\bar{T}(p,x)) \text{ and } h_{N-1,i}(p|x) = u(p,x),$$

the desired inequality is obtained. The other cases are considered similarly, and the proof is completed.

Proposition 2 yields Corollary 1, and Corollary 2 is easily obtained from Equation (7). Equation (10) yields $g_{N+1,i}(p) \geq 0$.

Proof of Proposition 4. **First we note that**

$$U_{N,i}(p) = \bigcup_{j=1}^{i-1} S_{N,i}(p) = \{ x \mid g_{N-1,i-1}(\bar{T}(p,x)) \leq u(p,x) \}.$$

If $x \in U_{N+1,i}(p)$, then $g_{N,i-1}(\bar{T}(p,x)) \leq u(p,x)$. Proposition 3 and Lemma 4 yield

$$g_{N,i-1}(\bar{T}(p,x)) \geq g_{N-1,i-1}(\bar{T}(p,x)).$$

Therefore we have

$$g_{N-1,i-1}(\bar{T}(p,x)) \leq u(p,x), \text{ i.e., } x \in U_{N,i}(p).$$

Proof of Proposition 5. Since $u(p,x) = w(x)$, if $x \in U_{N+1,i}(p)$, then $g_{N,i-1}(\bar{T}(p,x)) \leq w(x)$.

Proposition 1 and Lemma 4 yield

$$g_{N,i-1}(\bar{T}(p,x)) \geq g_{N,i-1}(\bar{T}(q,x)).$$

Therefore

$$g_{N,i-1}(\bar{T}(q,x)) \leq w(x), \text{ i.e., } x \in U_{N+1,i}(q).$$

The optimal policy and the total expected reward obtainable under this policy are embodied in the following theorem.

Theorem 1. Suppose a problem in state $(a_1,\ldots,a_N;p)$, then

1) $v_N(a_1,\ldots,a_N;p) = \sum_{i=1}^{N} a_i g_{N,i}(p)$.

2) When a realized value x of the random variables is observed, an optimal policy of the decision-maker is:

take the i-th action a_i and assign to the value x if $x \in S_{N,i}(p)$.

($i = 1,2,\ldots,N$)

Proof. We employ the induction principle on N. When N = 1, the problem $P_1(a_1;p)$ is obvious, and this theorem is valid since

$$g_{1,1}(p) = \sum_{i=1}^{\infty} p_i E_i u(p,X_i).$$

We assume this theorem for any value less than N. From the induction assumption, Equation (4) is rewritten as follows.

$$v_N(a_1,..,a_N;p|x) = \max_{1\le i\le N} \{ a_i u(p,x) + \sum_{j=1}^{i-1} a_j g_{N-1,j}(\bar{T}(p,x)) + \sum_{j=i+1}^{N} a_j g_{N-1,j-1}(\bar{T}(p,x)) \}. \quad (12)$$

Whenever $x \in S_{N,i}(p)$ ($i = 1, 2, \ldots, N$),

$$g_{N-1,i}(\bar{T}(p,x)) \le u(p,x) < g_{N-1,i-1}(\bar{T}(p,x))$$

from (8), and Proposition 2 yields the inequalities

$$g_{N-1,N-1}(\bar{T}(p,x)) \le \cdots \le g_{N-1,i}(\bar{T}(p,x)) \le u(p,x) < g_{N-1,i-1}(\bar{T}(p,x)) \le \cdots \le g_{N-1,1}(\bar{T}(p,x)). \quad (13)$$

The well known Hardy's theorem (see Section 1) yields, whenever $x \in S_{N,i}(p)$,

$$v_N(a_1,\ldots,a_N;p|x) = \sum_{j=1}^{i-1} a_j g_{N-1,j}(\bar{T}(p,x)) + a_i u(p,x) + \sum_{j=i+1}^{N} a_j g_{N-1,j-1}(\bar{T}(p,x)), \quad (14)$$

since $a_1 \ge \cdots \ge a_N$. Therefore if $x \in S_{N,i}(p)$, then an optimal decision

is to take the i-th action a_i and assign to the value x at this time.

Since $\bigcup_{j=1}^{N} S_{N,j}(p) = R_+$, Equation (3) is

$$v_N(a_1,\ldots,a_N;p) = \int_0^\infty v_N(a_1,\ldots,a_N;p\dagger x)dF_p(x)$$

$$= \sum_{j=1}^{N} \int_{S_{N,j}(p)} v_N(a_1,\ldots,a_N;p|x)dF_p(x). \qquad (15)$$

Substituting Equation (14) into (15) and rearranging the terms yield

$$v_N(a_1,\ldots,a_N;p) = \sum_{j=1}^{N} a_j \{ \int_{U_{N,j}(p)} g_{N-1,j-1}(\bar{T}(p,x))dF_p(x)$$

$$+ \int_{S_{N,j}(p)} u(p,x)dF_p(x) + \int_{L_{N,j}(p)} g_{N-1,j}(\bar{T}(p,x))dF_p(x) \}$$

$$= \sum_{j=1}^{N} a_j g_{N,j}(p).$$

The last equality is derived from Equation (10). Here we get the proof of this theorem.

Theorem 1 yields that $S_{N,i}(p)$ is the region where the i-th action a_i is taken under the optimal policy, and, therefore, the properties of the optimal policy are contained in Propositions 1 - 5.

In a sequential stochastic assignment problem with N actions treated in Derman et al. [3], Nakai [6] etc, the optimal policy is a critical number policy which is determined by the (N - 1) critical numbers, i.e., this policy is determined by a set of N intervals which is a partition of R_+. As for the problem treated here, the optimal policy is determined by the sets $S_{N,i}(p)$'s, which are not alwyas convex, i.e., the optimal policy is not always a critical number policy. Example 1 shows this fact.

Example 1. We treat this problem in a partially observable Markov chain with two states in the following manner; $p_{11} = p_{22} = \nu$, $w(p,x)$

$= u(p,x) = x$, X_i is an exponential with a mean $1/\lambda_i$ ($i = 1, 2$ and $f_i(x) = \lambda_i \exp(-\lambda_i x)$) and $\lambda_1 = 1.1$, $\lambda_2 = 3.1$.

In this problem S is considered as [0,1] and $p_1 \in [\ 0,1\ ]$. We assume that $0.5 \leq \nu = (2.1 - \eta)/(2 - \eta) \leq 1$ where $\eta = \lambda_2/\lambda_1 + \lambda_1/\lambda_2$.

When N = 1, a simple calculation yields $S_{1,1}(p) = R_+$,

$$g_{1,1}(p) = \int_0^\infty x dF_p(x) = p_1/\lambda_1 + p_2/\lambda_2,$$

and $\bar{T}_1(p,x) = (p_1 \nu f_1(x) + p_2 \nu' f_2(x))/(p_1 f_1(x) + p_2 f_2(x))$ for $p = (p_1, p_2)$ where $\nu' = 1 - \nu$.

When N = 2,

$$U_{2,2}(p) = R_+, \quad S_{2,0}(p) = \emptyset$$

and

$$S_{2,1}(p) = \{ x \mid g_{1,1}(\bar{T}(p,x)) \leq x \}$$

$$= \{ x \mid \bar{T}_1(p,x)/\lambda_1 + \bar{T}_2(p,x)/\lambda_2 \leq x \}.$$

Here we observe a set $S_{2,1}(p)$ with $p_1 = 1/(1+e)$, i.e., $p_2 = e/(1+e)$. Since

$$S_{2,1}(1/(1+e)) = \{ x \mid g(x) \leq x \}$$

with

$$g(x) = \{ (\nu+\nu'\lambda_1/\lambda_2)\exp(-\lambda_1 x) + (\nu+\nu'\lambda_2/\lambda_1)\exp(1-\lambda_2 x) \}$$
$$\times \{ 1/(\lambda_1 \exp(-\lambda_1 x) + \lambda_2 \exp(1-\lambda_2 x)) \}.$$

A simple calculation yields that $x = 0.5$ is an inflection point of $g(x)$ and $g(0,5) = 0.5$, $g'(0.5) > 1$, $g'(x) \geq 0$, $g(0) = 0$, $\lim_{x \to \infty} g(x) > 0$. Therefore the equation $g(x) = x$ has three roots $0 < \alpha_1 < 0.5 < \alpha_2$, and

$$S_{2,1}(1/(1+e)) = [\ \alpha_1, 0.5\ ] \cup [\ \alpha_2, \infty\ ),$$

i.e., $S_{2,1}(1/(1+e))$ is not a convex set.

Theorem 1 and Proposition 1 yield that the value $v_N(a_1,\ldots,a_N;p)$ is increasing in p for any values of $a_1,\ldots,a_N$. Moreover Theorem 1 and Proposition 2 yield that the conditional value $v_N(a_1,\ldots,a_N;p|x)$ defined by Equation (12) is increasing in x.

Finally, we consider a special case of this problem where $a_1 = \ldots = a_k = 1$ and $a_{k+1} = \ldots\ldots = a_N = 0$, i.e., a problem of optimal selections. Whenever the decision-maker observes a realization of a random variable associated with the current state of the process, he decides either to select this value or to reject it, and he can select k values out of N in order to maximize the total expected sum of k values selected. If k = 1, this is an optimal stopping problem which is formulated as a partially observable Markov decision process. We denote this problem as $P_{N,k}(p)$, and let the total expected reward obtainable under an optimal policy be $v_{N,k}(p)$.

Theorem 1 - 2) yields the following property for an optimal policy of $P_{N,k}(p)$.

Proposition 6. An optimal policy of the problem $P_{N,k}(p)$ is as follows. When a realization x is observed;

If $x \in U_{N,k+1}(p)$ then select this value,

or otherwise reject it.

In a sequential stochastic assignment problem treated in Derman,

Lieberman and Ross [3], etc, an optimal policy is determined by the critical numbers, and the following properties are obtained for these numbers. The critical number for an action associated with the k-th greatest value is increasing in N for any k ($k \leq N$). Proposition 4 corresponds to this fact; the region $U_{N,k+1}(p)$ consists of the values which can be selected by the decision-maker in the problem $P_{N,k}(p)$, and this region becomes smaller as N increases.

When w(p,x) is independent of p, i.e., u(p,x) is also independent of p, Proposition 5 yields that the region $U_{N,k+1}(p)$ becomes smaller as p increases. Theorem 1 - 1) yields the following proposition.

Proposition 7. $v_{N,k}(p)$ satisfies

$$v_{N,k}(p) = \sum_{j=1}^{k} g_{N,j}(p).$$

Proposition 7 yields that the total expected reward obtainable under an optimal policy for the problem $P_{N,k}(p)$ is a sum of k values $g_{N,i}(p)$ ($1 \leq i \leq k$). From Propositions 1, 3 and 7, $v_{N,k}(p)$ is increasing in p and N. (N = 1, 2, ... and $p \in S$) Proposition 7 yields

$$g_{N,k}(p) = v_{N,k}(p) - v_{N,k-1}(p).$$

Thus $g_{N,k}(p)$ means a difference between two values, $v_{N,k}(p)$ and $v_{N,k-1}(p)$, In this problem of optimal selections, $g_{N,k}(p)$ is considered as an extra profit of another "select" action to the problem $P_{N,k-1}(p)$: if the decision-maker can take another "select" action in the problem $P_{N,k-1}(p)$, he can expect to obtain $g_{N,k}(p)$ from this additional action. From Propositions 1 - 3, $g_{N,k}(p)$, which is decreasing in k and increasing in N

and p ($1 \leq k \leq N$ and $p \in S$), is considered as a worth of the k-th "select" action in the problem $P_{N,k}(p)$.

As concerns a sequential stochastic assignment problem treated here, $g_{N,k}(p)$ is an expected quantity by an action associated with the k-th greatest value under an optimal policy in the problem $P_N(a_1,\ldots,a_N;p)$. ($1 \leq k \leq N$) In this problem the k-th action has own value a_k and the decision-maker expects a quantity $g_{N,k}(p)$ by this action, and, therefore, the expected reward obtainable by this action under an optimal policy is $a_k g_{N,k}(p)$.

The number of jobs is, however, known in most of the sequential stochastic assignment problems. It is natural to study the situation where this number is not known in advance but is a random variable. A problem of this type is studied in Nakai [9], [10], where a problem with a knowledge of a prior distribution about the actual number of jobs is treated in the following manner. Suppose there are N jobs, and the number N is assumed to be a random variable whose distribution is given beforehand. Regarding the number of remaining jobs, all information is summarized by a probability distribution $q = (q_0,\ldots,q_M)$ on the set $\{ 0,1,\ldots,M \}$. Consider that N jobs are labelled 1, 2,...., N. Let Z_j be an arrival time of the job labelled j, and we assume that $Z_1,\ldots,Z_N$ are independently and identically distributed exponential random variables with a known mean $1/\lambda$, i.e.,

$$P(Z_j \leq t) = 1 - \exp(-\lambda t). \quad (j = 1,2,\ldots,N)$$

The information regarding the number of remaining jobs is updated in a Bayesian mannner as the successive jobs are observed.

Let X_j, $j = 1,2,\ldots,N$, be a value of the job labelled j, and the X's

are independently and identically distributed random variables with a common cumulative distribution function F(x) which has a finite mean.

Under the above conditions, a sequential stochastic assignment problem of this case is characterized by the following four things. 1) The planning time period T remaining at the last job offer. 2) The passage time t since the last job offer. 3) Information q about the number N of remaining jobs at the last job offer. Here we assume that $P (N \leq M \mid q) = 1$ for a given constant M. 4) The set $\{ a_1,\ldots,a_n \}$ of available actions. Similarly to the problem treated in [3], it is assumed that M = n without loss of generality. Under the above conditions, we will consider the $(a_1,\ldots,a_n;T,t,q)$ as a state variable. Whenever a job arrives, a decision based on $(a_1,\ldots,a_n;T,t,q)$ is made by the decision-maker, and, therefore, we treat this problem by choosing these points of time, so as to exploit the lack of memory of the exponential distribution.

Whenever a new job arrives at time t since the last job offer with a realization x of the random variable x, the decision-maker updates information about the number of remaining jobs and selects one of n available actions. The objective of this problem is to maximize the total expected reward. Similarly to the problem treated here, each action is used only once, and this problem stops whenever there is no action or T = 0. An optimal policy and the total expected reward obtainable under this policy are discussed in [10].

References

[1] Albright, S.C. (1977). A Bayesian Approach to a Generalized House Selling Problem. Management Science. 24 432-440.

[2] DeGroot, M.H. (1970). Optimal Statistical Decisions. McGraw-Hill,

New York.

[3] Derman, C., G.J. Lieberman and S.M. Ross. (1972). A Sequential Stochastic Assignment Problem. Management Science. 18 349-355.

[4] Hardy, G.H., J.E. Littlewood and G. Polya. (1934). Inequalities. Cambridge, England.

[5] Monahan, G. (1980). Optimal Stopping in a Partially Observable Markov Chain with Costly Information. Operations Research. 28 1319-1334.

[6] Nakai, T. (1982). A Time Sequential Game Related to the Sequential Stochastic Assignment Problem. Journal of Operations Research Society of Japan. 25 129-138.

[7] Nakai, T. (1984). The Problem of Optimal Stopping in a Partially Observable Markov Chain. Accepted for publication in Journal of Optimization Theory and Applications.

[8] Nakai, T. (1984). A Sequential Stochastic Assignment Problem in a Stationary Markov Chain. Submitted to Mathematica Japonica.

[9] Nakai, T. "Optionの数が未知である Sequential Stochastic Assignment Problem" 日本数学会1981年度秋季総合分科会.

[10] Nakai, T. Optimal Assignment for a Random Sequence with an Unknown Number of Jobs. Submitted to Journal of Opetations Research Society of Japan, on December, 1982.

[11] Ross, S.M. (1970). Applied Probability with Optimization Applications. Holden-Day, San Fransisco.

[12] Ross, S.M. (1971). Quality Control under Markovian Deterioration. Management Science. 17 587-596.

[13] Ross, S.M. (1983). Stochastic Processes. John-Wiley & Sons, New York.

[14] Sakaguchi, M. (1972). A Sequential Assignment Problem for Randomly Arriving Jobs. Reports on Statistical Applications Research, Union of Japanese Scientists and Engineers. 19 15-25.